

CASE STUDY · HR AUTOMATION

Six documents in, one payroll row out.

Scanned employee paperwork read by AI, cross-checked against itself, and written into the client's own payroll template — every value traceable to the page it came from.

TIME RETURNED

~12 hrs

of manual capture removed from every 40-candidate intake — the batch size the client runs

COST AVOIDED

~\$2,700

a year against an always-on server, measured from the platform's own usage logs

98%

of fields filled with no human touch in the run shown here

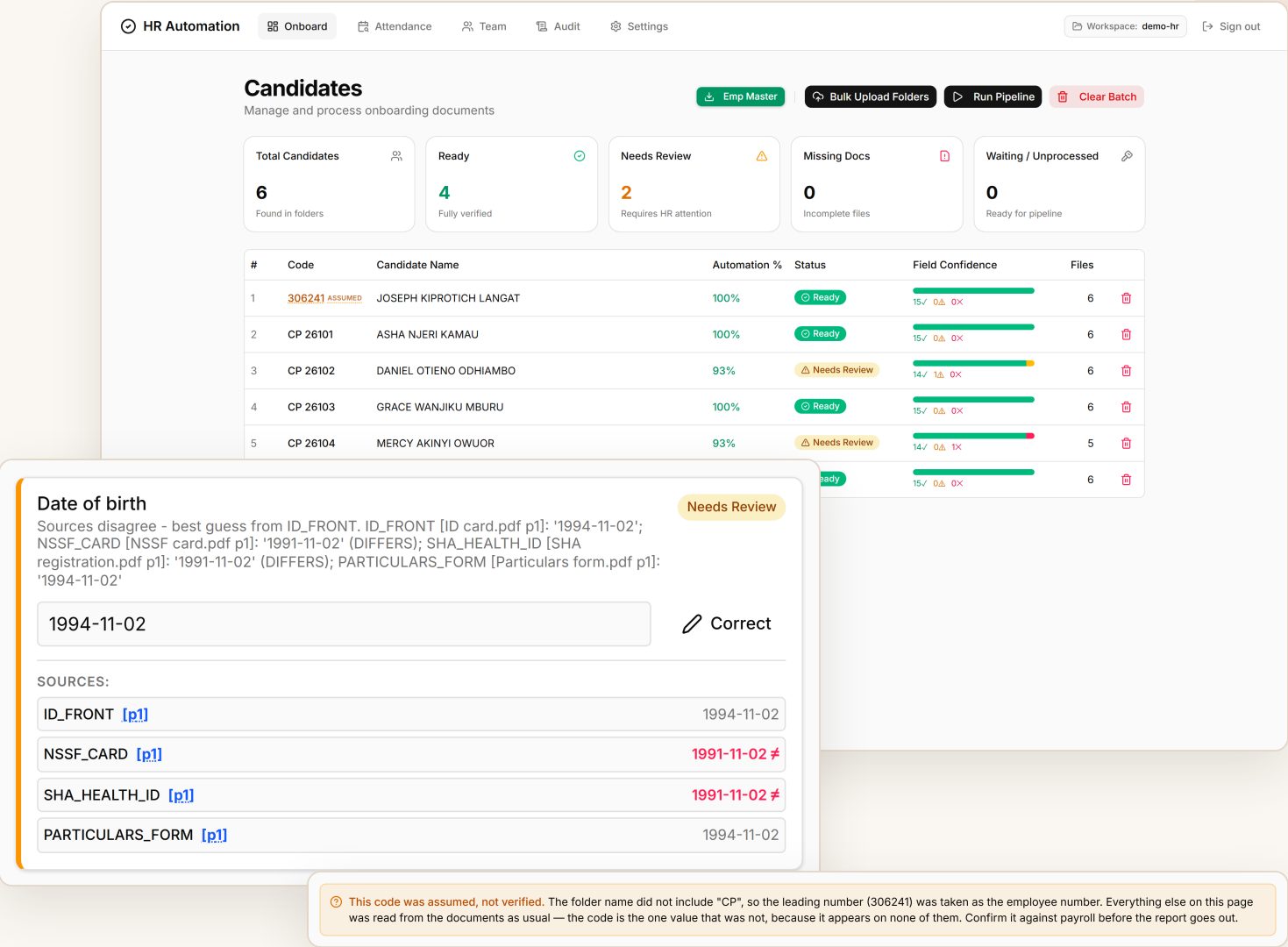
54

columns per person, filled and derived automatically

\$0

running cost between intakes — it scales to zero

Time saving from HR's own estimate of 15–20 minutes per candidate to read six scans, retype ~15 fields and reconcile values that appear on more than one document. Screens throughout show the delivered product running on invented records.



HOW IT WORKS

Drop the folders in. Everything after that is automatic — until a judgement call is needed.

01

Folders in

Dragged in by the hundred and staged. The folder names the candidate, at any nesting depth.

02

Pages rendered

Every PDF page and photograph rendered to an image and cached, so a re-run costs nothing.

03

Each page read

A vision model classifies the page — ID, KRA, NSSF, SHA, contract — and returns only its fields.

04

Cross-checked

The same field from different documents is compared and format-validated, then agreed or flagged.

05

Routed

Confident values are written. Anything ambiguous goes to a person with the evidence attached.

06

Sheet out

The payroll template filled, stamped with who exported it, confirmations logged.

Add candidates

×

Drag candidate folders in. Add as many times as you like — nothing uploads until you press the button at the bottom.

⬆

Drag candidate folders here

Ctrl-click as many as you like and drop them together, or drop the folder that **contains** them to add every one inside. Each candidate needs its own folder, named CP 26002 NAME or 306252 NAME. PDFs and images (JPG, PNG) both work.

Drop as many times as you like — they add up. Nothing uploads until you press the button below.

Close

Upload and process documents

Staged, not uploaded on arrival. Drop thirty, then another hundred — nothing leaves the browser until the button is pressed, and nothing is discarded in silence. Drag-and-drop only, deliberately: a browser folder picker can return just one directory, so a Browse button invites an attempt that cannot work.

Classified by content, not filename

Six separate files, one combined PDF, or a phone photo of a single certificate — all arrive at the same place. A badly named file changes nothing.

Work is never repeated

Readings are cached per page. Adding one late document to a batch of forty re-reads one page, not four hundred — and every page read is a metered call.

Total Candidates

6

Found in folders

Ready

4

Fully verified

Needs Review

2

Requires HR attention

Missing Docs

0

Incomplete files

Waiting / Unprocessed

0

Ready for pipeline

Batch totals as the run progresses — the numbers HR watches.

PIPELINE

02

Which people need a human, and which do not.

DELIVERED SCREEN

#	Code	Candidate Name	Automation %	Status	Field Confidence	Files
1	306241 ASSUMED	JOSEPH KIPROTICH LANGAT	100%	Ready	<div><div></div></div> 15✓ 0△ 0×	6
2	CP 26101	ASHA NJERI KAMAU	100%	Ready	<div><div></div></div> 15✓ 0△ 0×	6
3	CP 26102	DANIEL OTIENO ODHIAMBO	93%	Needs Review	<div><div></div></div> 14✓ 1△ 0×	6
4	CP 26103	GRACE WANJIKU MBURU	100%	Ready	<div><div></div></div> 15✓ 0△ 0×	6
5	CP 26104	MERCY AKINYI OWUOR	93%	Needs Review	<div><div></div></div> 14✓ 0△ 1×	5
6	CP 26105	BRIAN MUTUA KILONZO	100%	Ready	<div><div></div></div> 15✓ 0△ 0×	6

Automation %

How much of that person's row was filled without anyone touching it.

Field confidence

Confirmed, needs review, missing — so a 93% row shows which single field is the problem.

Status is derived

Ready means every field either agreed across documents or was confirmed by a named person.

Amber means assumed

A code taken from a folder name rather than read from a document is never shown as verified.

Six invented candidates after one pipeline run: four clear, two holding a single field each for a person to decide. The exceptions are the work — the rest is already done.

Every value carries its receipts — and contradictions stop the line.

DELIVERED SCREENS

Email address

3 sources agree

Confirmed

asha.njeri@example.com

SOURCES:

KRA_CERT [p1]asha.njeri@example.com

CONTRACT_LETTER [p1]asha.njeri@example.com

PARTICULARS_FORM [p1]asha.njeri@example.com

Three documents agree. Each source is named, with the page it was read from — [p1] opens that rendered page, so a value can be checked against the scan in one click. Written automatically.

Date of birth

Sources disagree - best guess from ID_FRONT. ID_FRONT [ID card.pdf p1]: '1994-11-02'; NSSF_CARD [NSSF card.pdf p1]: '1991-11-02' (DIFFERS); SHA_HEALTH_ID [SHA registration.pdf p1]: '1991-11-02' (DIFFERS); PARTICULARS_FORM [Particulars form.pdf p1]: '1994-11-02'

Needs Review

1994-11-02

Correct

SOURCES:

ID_FRONT [p1]1994-11-02

NSSF_CARD [p1]1991-11-02 ≠

SHA_HEALTH_ID [p1]1991-11-02 ≠

PARTICULARS_FORM [p1]1994-11-02

Two documents disagree. The NSSF card and SHA record read 1991; the ID and particulars form read 1994. Two against two — so nothing is written until a person confirms.

Majority agrees → written

Enough independent documents say the same thing, and it passes the format rule for that field.

One dissenter → outvoted, and noted

A lone mismatch against a firm majority is overruled — but the disagreement is recorded on the field rather than discarded.

No majority → a person decides

The pipeline refuses to choose, shows its best guess and every reading behind it, and waits.

CROSS-VALIDATION

04

Marking a guess as a guess.

DELIVERED SCREENS

 **This code was assumed, not verified.** The folder name did not include "CP", so the leading number (306241) was taken as the employee number. Everything else on this page was read from the documents as usual — the code is the one value that was not, because it appears on none of them. Confirm it against payroll before the report goes out.

Some intakes name folders by payroll number instead of the expected code. Those folders used to vanish from the batch without a word — indistinguishable, from the outside, from a failed upload.

#	Code	Candidate Name	Automation %	Status	Field Confidence	Files
1	306241 ASSUMED	JOSEPH KIPROTICH LANGAT	100%	 Ready	<div><div></div><div>15✓ 0Δ 0×</div></div>	6 
2	CP 26101	ASHA NJERI KAMAU	100%	 Ready	<div><div></div><div>15✓ 0Δ 0×</div></div>	6 

The same amber marking follows the code onto the batch list and the staging dialog — impossible to mistake for one that was read off a document.

Accepted, and kept as written

Inventing a **CP** prefix would put a code nobody uses into the payroll column — worse than the original problem.

But flagged as assumed

It is the one value never read from a document. Verified across a real sample: the code appears in the text of **none** of the paperwork.

Photographs count as documents

A phone picture of a certificate is read like any scan — camera rotation corrected, size capped — instead of the candidate appearing to be missing paperwork they supplied.

Five digits minimum

A bare number needs five or more digits to be believed. Four digits is a year far more often than an employee number — without that floor a subfolder named **2024 scans** quietly became a candidate of its own.

The client's own template, filled in — and a log of who confirmed what.

DELIVERED OUTPUT

SYSTEM NO	PREFIX	KENYA EE CODE	EE FULL NAMES	SURNAME	FIRST NAMES	EE ID NO.	DATE OF BIRTH	ENGAGEMENT
KENM	126	306241	JOSEPH KIPROTICH LANGAT	LANGAT	JOSEPH KIPROTICH	37714002	2000-01-30	2026-03-02
KENM	126	CP 26101	ASHA NJERI KAMAU	KAMAU	ASHA NJERI	34118276	1996-04-18	2026-03-02
KENM	126	CP 26102	DANIEL OTIENO ODHIAMBO	ODHIAMBO	DANIEL OTIENO	31904553	1994-11-02	2026-03-02

Values exactly as the pipeline wrote them during the run shown throughout — 9 of the 54 columns. Surname and first names are split from the full name; initials, tax and leave dates computed. The amber row is the assumed code, carried through unchanged.

When	Who	Action	Workspace	Details
24 Aug 2026, 17:59:10	demo-hr	Signed in	demo-hr	—
24 Aug 2026, 17:53:00	demo-hr	Signed in	demo-hr	—
24 Aug 2026, 17:51:37	demo-hr	Signed in	demo-hr	—
24 Aug 2026, 17:50:28	demo-hr	Downloaded an export	demo-hr	Emp_Master_FILLED_20260824_175027-demo-hr.xlsx (DL-C7E58BCB)
24 Aug 2026, 17:47:37	demo-hr	Signed in	demo-hr	—
24 Aug 2026, 17:47:27	demo-hr	Signed in	demo-hr	—
24 Aug 2026, 17:42:53	demo-hr	Signed in	demo-hr	—
24 Aug 2026, 17:40:35	demo-hr	Started the pipeline	demo-hr	all candidates
24 Aug 2026, 17:40:35	demo-hr	Signed in	demo-hr	—

“Who confirmed these details?” — the client's question, asked plainly. The log records that a named person confirmed a named field, **never the value**: an audit trail should not become a second database of ID numbers. It is kept outside every workspace, so deleting an account cannot erase it.

Team

Create an account for each team member. They sign in with the workspace name and password you give them, and can only see their own candidates.

New account

The workspace name becomes their sign-in name — a first name works well. Letters, numbers and dashes only.

Workspace name

e.g. caroline

Password

At least 8 characters

Create

The password is shown as you type so you can pass it on accurately. It is stored hashed and cannot be read back afterwards — if it's forgotten, reset it here.

Accounts

Workspace

Created

Last sign-in

Actions

demo-hr

admin

24 Aug 2026, 17:40

24 Aug 2026, 17:59

Reset

One admin creates a workspace per team member. Each person sees only their own candidates, enforced in the storage layer — so no endpoint can leak another workspace's batch, whatever it does.

Costs nothing while nobody is using it.

Onboarding is bursty — a busy week per intake, then quiet. The architecture was matched to that shape rather than to a server left running.

- 01

Scales to zero
No container runs between batches. Billing follows use, not the calendar.
- 02

Heavy work runs as a job
Extraction runs on a cheaper billing model than a web service, and a long batch cannot time out a request.
- 03

Pages cached, runs resumable
The metered cost is per page read. Re-running a batch re-reads nothing it has already seen.
- 04

Eight candidates at a time
Concurrency tuned and capped — fast on a big intake, without a runaway bill.
- 05

Rotation, retries and resume
Sideways scans can be rotated in place; an interrupted run continues instead of starting over.

MEASURED, NOT ESTIMATED

~\$2.7k

a year avoided against an always-on equivalent, from the platform's own usage logs and the provider's published rates

3.9 hrs

of processor time consumed by every batch in the measured month — against the 730 hours an always-on server is billed for

What was delivered

A working system in production use: bulk intake, page-level document AI, cross-validation with human routing, per-member access, an audit trail, rotation and resume, and export into the client's existing payroll template — deployed, and handed over with the failure modes written down.

The bigger number is the staff time. About 12 hours per 40-candidate intake, on HR's own estimate of 15–20 minutes a person. The infrastructure figure above is the one that can be measured to the cent.